

Conference Abstract

Data Migration from One Database to Another: Nervous breakdown of a database manager!

Henry Engledow [‡]

[‡] Meise Botanic Garden, Meise, Belgium

Corresponding author: Henry Engledow (henry.engledow@plantentuinmeise.be)

Received: 14 Jun 2019 | Published: 21 Jun 2019

Citation: Engledow H (2019) Data Migration from One Database to Another: Nervous breakdown of a database manager! Biodiversity Information Science and Standards 3: e37302. <https://doi.org/10.3897/biss.3.37302>

Abstract

Migrating from one database to another is always accompanied by challenges. Recently Meise Botanic Garden has migrated its Living Collections data from *LivCol*, a bespoke database, to *BG-BASE*, a commercial product. Differences in database structure, degree of atomisation and field definition increase the complexity of such a transfer of data. The greater the number of fields used in the original data source, the greater the number of problems there are to resolve. Living collections are often centered around the 'accession information' of the living material, but the way one does this and the philosophy behind this may differ. The different approach to accessioning material in *LivCol* and *BG-BASE* affected the structure of the data model in each program. The *LivCol* approach was not as strictly defined as the *BG-BASE* approach e.g. new generations derived from existing accessions in *LivCol* retained the same accession number despite being not genetically identical (of seed origin), whereas in *BG-BASE* a new accession number would be generated with reference to the parent accession. In the data transfer *LivCol* accession number where grouped by accession number and garden location, and the inter-generation information combined in a single record in *BG-BASE* (this is not ideally in accordance with the *BG-BASE* concept, the alternative was to create 'artificial' new accession numbers but this would have complicated matters more both from a data and management point of view).

The use of standards would greatly improve data transfer, and indeed many standards have been adopted by both the above mentioned databases. However, it soon became

evident that there are multiple standards for a single topic e.g. for information concerning conservation status: NatureServe Global Conservation Status Ranks; Fish & Wildlife conservation category; International Union for Conservation of Nature (IUCN) - old and new codes (plus version); Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES); etc. Also, some values used are region specific and do not translate well to 'global' standards e.g. in Belgium there are three principal regions controlling conservation status (Brussels, Flanders, Wallonia) and they differ in their approach and definitions, all these need to be taken into account as there are legal implications - this was done by finding close matches in IUCN (New) codes and combining them with 'non-standard' World Geographical Scheme for Recording Plant Distributions (WGSRPD). The latter TDWG standard is out of date and in many circumstances not sufficiently atomised to be of practical use. There were also certain fields that would benefit from having standards, but are at present absent e.g. invasiveness - *BG-BASE* uses Cronk and Fuller (1995) whereas *LivCol* uses AlterIAS (<http://www.alterias.be/>), Belgian Forum on Invasive Species (<http://ias.biodiversity.be>) and Lambinon et al. (1992). The above problems will be discussed and their impact on mapping the data. Decisions had to be made with respect to 'best fit' solutions. The latter lead to the loss of information or a slight variation in its interpretation, examples will be given to highlight these aspects. As the structures of the databases differed, sometimes assumptions had to be made, this too will be illustrated. These changes were reasonable, but represent an interpretation of the original data and therefore not strictly the same.

Databases are crucial to the management of Living Collections and the research done on them. This talk will look at the lessons learned during the data transfer and the problems associated with mapping (decisions, assumptions and standards). Databases are 'living' entities that need to grow, adapt, be maintained and regularly updated to new developments in technology. Databases are not seen as new or innovative by funding bodies and are often left to struggle along in suboptimal conditions. If we want data quality to improve and increase interoperability between systems, maybe we should start at the point where data is entered.

Keywords

databases, data quality, interoperability, standards

Presenting author

Henry Engledow

Presented at

Biodiversity_Next 2019

References

- Cronk QC, Fuller JL (1995) Plant Invaders. The threat to natural ecosystems. Chapman & Hall, London, xiv, 241 pp. [ISBN 0-412-48380-7]
- Lambinon J, Langhe J-D, Delvosalle L, D'Hose R, Geerinck DJ, Lebeau J, Schumacker R, Vannerom H, Rammeloo J, Duvigneaud J (1992) Nouvelle flore de la Belgique, du Grand-Duché de Luxembourg, du nord de la France et des régions voisines (Ptéridophytes et Spermatophytes). 4. Jardin Botanique National de Belgique, Meise, cxx, 1092 pp. [ISBN 9072619072]